

WHAT IS CLAIMED IS:

1. A floating point unit comprising:

5 a multiplier configured to generate a first result;

an approximation circuit coupled to receive the first result and configured to
generate an approximation of a difference of the first result from a
constant; and

10

a control circuit coupled to the multiplier and the approximation circuit, wherein
the control circuit is configured to approximate a function specified by a
floating point instruction provided to the floating point unit for execution
using an approximation algorithm, the approximation algorithm
15 comprising at least two iterations through the multiplier and optionally the
approximation circuit, wherein the control circuit is configured to correct
the approximation from the approximation circuit from a first iteration of
the approximation algorithm during a second iteration of the
approximation algorithm by supplying a correction vector to the multiplier
20 during the second iteration;

20

wherein the multiplier is configured to incorporate the correction vector into the
first result during the second iteration.

25 2. The floating point unit as recited in claim 1 wherein the multiplier is coupled to
receive at least two operands, and wherein the first result of the multiplier comprises a
sum of the correction vector and a product of the operands.

3. The floating point unit as recited in claim 2 wherein the approximation is less than the

difference by one unit in the least significant bit of the approximation.

4. The floating point unit as recited in claim 3 wherein the correction vector comprises binary zeros except for a binary one in a location corresponding to the least significant bit of the approximation.

5. The floating point unit as recited in claim 4 wherein the approximation includes one or more extra precision bits in addition to a precision specified by a floating point standard, and wherein the least significant bit of the approximation is a least significant bit of the extra precision bits.

6. The floating point unit as recited in claim 1 wherein the approximation circuit is configured to generate the approximation by inverting the first result and optionally shifting the inverted first result.

7. The floating point unit as recited in claim 1 wherein the approximation algorithm is recursive.

8. The floating point unit as recited in claim 7 wherein the approximation algorithm is the Newton-Raphson algorithm.

9. The floating point unit as recited in claim 1 wherein the correction vector is input on a rounding vector input of the multiplier, and wherein the control circuit is configured to transmit a rounding vector on the rounding vector input in response to an operation for which rounding is specified.

10. A method comprising:

approximating a function using an approximation algorithm comprising at least

two iterations of: (i) a multiplication in a multiplier, and (ii) a difference approximation between a constant and a result of the multiplication, wherein the difference approximation is less than a difference between the constant and the result by one unit in the least significant bit of the difference approximation;

during a first of the iterations, generating the difference approximation; and

during a second of the iterations, supplying a correction vector to the multiplier, the multiplier incorporating the correction vector in the multiplication, the correction vector correcting the difference approximation.

11. The method as recited in claim 10 wherein a result of the multiplication comprises a sum of the correction vector and a product of at least two operands of the multiplication.

12. The method as recited in claim 11 wherein the correction vector comprises binary zeros except for a binary one in a location corresponding to the least significant bit of the difference approximation.

13. The method as recited in claim 12 wherein the difference approximation includes one or more extra precision bits in addition to a precision specified by a floating point standard, and wherein the least significant bit approximation is a least significant bit of the extra precision bits.

14. The method as recited in claim 10 wherein the approximation algorithm is recursive.

15. The method as recited in claim 14 wherein the approximation algorithm is the Newton-Raphson algorithm.

16. A carrier medium comprising one or more data structures representing a floating point unit, the floating point unit comprising:

a multiplier configured to generate a first result;

5

an approximation circuit coupled to receive the first result and configured to generate an approximation of a difference of the first result from a constant; and

10

a control circuit coupled to the multiplier and the approximation circuit, wherein the control circuit is configured to approximate a function specified by a floating point instruction provided to the floating point unit for execution using an approximation algorithm, the approximation algorithm comprising at least two iterations through the multiplier and optionally the approximation circuit, wherein the control circuit is configured to correct the approximation from the approximation circuit from a first iteration of the approximation algorithm during a second iteration of the approximation algorithm by supplying a correction vector to the multiplier during the second iteration;

15

20

wherein the multiplier is configured to incorporate the correction vector into the first result during the second iteration.

17. The carrier medium as recited in claim 16 wherein the multiplier is coupled to receive at least two operands, and wherein the first result of the multiplier comprises a sum of the correction vector and a product of the operands.

25

18. The carrier medium as recited in claim 17 wherein the approximation is less than the difference by one unit in the least significant bit of the approximation.

19. The carrier medium as recited in claim 18 wherein the correction vector comprises binary zeros except for a binary one in a location corresponding to the least significant bit of the approximation.

5

20. The carrier medium as recited in claim 19 wherein the approximation includes one or more extra precision bits in addition to a precision specified by a floating point standard, and wherein the least significant bit of the approximation is a least significant bit of the extra precision bits.

10

21. The carrier medium as recited in claim 16 wherein the approximation circuit is configured to generate the approximation by inverting the first result and optionally shifting the inverted first result.

15

22. The carrier medium as recited in claim 16 wherein the approximation algorithm is recursive.

23. The carrier medium as recited in claim 22 wherein the approximation algorithm is the Newton-Raphson algorithm.